Biometrical Journal **48** (2006) 1, 54–66 DOI: 10.1002/bimj.200510201

Standardized Risk and Description of Results from Multivariable Modeling of a Binary Response

Bong-Rae Kim[*,1], **Randy L. Carter**[2], **P. V. Rao**[1], **Mario Ariet**[3], and **Michael B. Resnick**[3]

[1] Department of Statistics, University of Florida, Gainesville, FL 32611, USA
[2] Department of Biostatistics, University at Buffalo, Buffalo, NY 14214, USA
[3] Maternal Child Health and Education Research and Data Center, University of Florida, Gainesville, FL 32608, USA

Received 28 July 2005, revised 31 October 2005, accepted 23 November 2005

Summary

Descriptions of significant associations found from a logistic regression analysis typically are based on adjusted odds ratios. Unfortunately, odds ratios provide no information about the prevalence of response. In this paper, we justify and recommend using standardized risks, i.e., standardized probabilities, which do provide information about prevalence, in addition to adjusted odds ratios, for pairwise comparisons of the levels of a significant factor. We illustrate the advantages of generally reporting standardized risk estimates, in the context of assessing the effect of blood lead levels during the preschool years on occurrence of academic problems in kindergarten. Results are more meaningfully interpreted when accompanied by standardized risk estimates.

Key words: Logistic regression; *Post hoc* comparisons; Adjusted odds ratios; Standardized proportions; Standardized attributable risk; Standardized risk; Standardized risk difference; Binary response models; Lead poisoning; Academic problems.

1 Introduction

Neison (1844) generally is credited with first proposing the method of direct standardization to control for potentially confounding factors when comparing populations with respect to disease or death rates (Lilienfeld, 1979). Standardized rates and risks are used today as descriptive statistics, and methods of standardization still are presented in books on modern epidemiology (Rothman, 1998; Bhopal, 2002). Standardization is viewed by many, however, as the "humble forerunner" of modern statistical modeling (Fletcher et al., 1982). With the development of modeling methods, such as logistic regression, the method of standardization lost favor. Because of their mathematical relationship with model parameters, adjusted odds ratios (AORs) gained popularity for *post hoc* pairwise comparisons of significant effects with the increased use of logistic regression for primary analysis. Nevertheless, standardized risks and risk differences hold an advantage over AORs. That is, they convey information about the prevalence of the response in addition to information about the magnitude of association. Differences between category specific standardized risks (i.e., standardized attributable risks (SARs)), therefore, are more naturally appealing and meaningfully interpreted than the adjusted odds ratios that result from logistic regression analysis.

The need for meaningful interpretations probably explains why estimators of risk and risk differences are still used by some instead of estimation on the log odds scale. The concept of attributable

[*] Corresponding author: e-mail: bkim@stat.ufl.edu, Phone: +01 352 392 3321 Ext. 230

risk, in the form of various measures (all functions of a risk difference), was used increasingly in the medical literature from 1966 to 1996 (Uter and Pfahlberg, 1999). Several articles in the statistical literature acknowledged the advantage of interpretability held by standardized risk differences and the mathematical advantage of log odds ratios in the context of logistic models (Greenland and Holland, 1991; Greenland, 1991; Flanders and Rhodes, 1987). As a result the authors suggested estimators of standardized differences based on adjusted odds ratios or logistic model parameter estimators.

In logistic regression analysis, the popularity of adjusted odds ratios as measures of association and descriptions of significant effects stems from the fact that *post hoc* pairwise comparison hypotheses about the effects of a covariate on the linear predictor are easily tested as hypotheses about odds ratios. Most importantly, there is a known equivalence of the null hypothesis that an adjusted odds ratio is one and the null hypothesis that the corresponding effect on the linear predictor is zero. Furthermore, the magnitude of violation of either null hypothesis is a monotonic function of the other. Thus, adjusted odds ratios are justified descriptors of associations between categorical covariates and a binary response.

Unfortunately, while adjusted odds ratios measure the magnitude of effects well, they provide no sense of the prevalence or difference in prevalence among levels of the explanatory factor of interest. Standardized risks have an advantage over adjusted odds ratios in this regard, provided they are calculated from representative standard populations. This advantage is the likely cause of renewed interest in describing effects on the risk scale (i.e., probability scale) instead of the odds scale. Recent versions of the SAS and STATA software packages, for example, have added functionality for calculating estimates of standardized probabilitites (PROC SURVEYREG in SAS and the SVY TABULATE command in STATA). STATA also has built in support in its binary regression routines for estimating adjusted probabilities. The adjust command, used with the logit command, for example, produces model-based estimates of standardized risks. (As an aside, it is unfortunate that the computations are limited to a uniform standard distribution, which often does not adequately reflect the population under study. Nevertheless, the standardized risk differences that are calculated from the resulting adjusted probability estimates do measure effects adjusted for confounding factors.) The increased popularity of standardized risk differences, while welcomed, is not yet backed by a mathematical justification similar to that supporting AORs.

The primary purpose of this paper is to justify the use of SARs as descriptors of significant associations observed in a logistic regression analysis. Secondarily, we extend the justification to more general models for binary responses with probability functions that, like logistic models, are monotonic in a linear predictor.

We start with a brief review of the logistic model in Section 2. Justification of the *post hoc* descriptors, AOR and SAR, of effects found in the logistic regression is presented in Section 3 along with the methods for computing standard errors and confidence intervals for standardized risks and SARs. The use of descriptors based on standardized risks is illustrated in the context of a research study described in Section 4. The similar justification in generalized linear model analysis is presented in Section 5. In Section 6, we present conclusions and recommendations.

2 Logistic Model

Logistic regression models are commonly used to investigate the relationship between a binary response and a set of explanatory variables, not only because of appealing mathematical properties, but also because of the simple interpretation of model parameters as the logarithm of the odds ratios (McCullagh and Nelder, 1989). For binary response models, the response variable, Y, can take only one of two possible values, denoted for convenience by 0 and 1. Let X be a categorical explanatory variable of interest with k categories, and let $\mathbf{Z}$ denote the vector of c other covariates. Also assume that $\mathbf{Z} = z_i$ is the i-th combination of possible values of the components of $\mathbf{Z}$, where $z_i = (z_{i1}, \ldots, z_{ic})^T$, $i = 1, 2, \ldots, s$, and s is the number of strata defined by all combinations of the $\mathbf{Z}$ covariates.

Let $\pi_{ij} = \Pr(Y = 1 \mid \mathbf{Z} = z_i, X = j)$ be the probability of the response of interest for the i-th stratum within the j-th level of X, $j = 1, 2, \ldots, k$. Following common practice, we assume that the dependence of π_{ij} on the explanatory variables occurs through the linear predictor for i-th stratum within the j-th level of X, defined by

$$lp_{ij} = \alpha + \beta_j + \sum_{g=1}^{c} \delta_g z_{ig} \,, \tag{1}$$

where α, β_j, and δ_g are unknown regression coefficients. The logistic model, then, has the form

$$\pi_{ij} = \Pr(Y = 1 \mid \mathbf{Z} = z_i, X = j) = \frac{\exp\{lp_{ij}\}}{1 + \exp\{lp_{ij}\}} \,. \tag{2}$$

3 Justification for AOR and SAR in Logistic Models

In this section, we review the justification of adjusted odds and odds ratios as *post hoc* descriptors of overall effects in a logistic regression. We then show that standardized risks and standardized attributable risks are similarly justified. Three methods for computing standard errors of standardized risk estimators and their differences are presented in Section 3.4.

3.1 Odds and AOR

Given the logistic model in (1) and (2), we have $\beta_l - \beta_m = \text{logit}(\pi_{il}) - \text{logit}(\pi_{im})$. So the logistic model expresses the differences in the effects of two given levels of X as an appropriate difference between the probabilities of response measured on the logit scale. Thus, to understand the practical meaning of an analysis based on the logistic model, it is important to appreciate the meaning of differences on the logit scale. As pointed out by Cox and Snell (1989), differences on the logit scale can be easily grasped by thinking from the beginning not in terms of the probability of the response of interest ($Y = 1$) but rather in terms of the odds of $Y = 1$ against $Y = 0$, namely $\pi_{ij}/(1 - \pi_{ij})$. The exponential of a logit difference is then a ratio of odds.

The adjusted odds ratio for comparing the probability of a response of interest given $X = l$ with that given $X = m$ within the i-th stratum is defined by

$$\text{AOR}_{i(l,m)} = \frac{\pi_{il}/(1 - \pi_{il})}{\pi_{im}/(1 - \pi_{im})} = \exp\{lp_{il} - lp_{im}\} \,. \tag{3}$$

Given lp_{ij} as defined in (1), this adjusted odds ratio can be written as

$$\text{AOR}_{i(l,m)} = \exp\{\beta_l - \beta_m\} \tag{4}$$

and is the same for all $i = 1, 2, \ldots, s$. Henceforth, $\text{AOR}_{l,m}$ will denote the common odds ratio, which is referred to as the adjusted ratio for comparing level l to level m of the categorical covariate of interest, X.

The $\text{AOR}_{l,m}$, $1 \le l \neq m \le k$, and their $(1 - \alpha) \times 100\%$ confidence intervals typically are used in *post hoc* pairwise comparisons to more fully understand the significant overall effect of X on Y (see, e.g., Freeman, 1987; Hosmer and Lemeshow, 1989).

The use of AORs as *post hoc* descriptors of a significant effect of X in logistic regression is justified by:

1. The null hypothesis $H_{01} : \beta_l - \beta_m = 0$ is equivalent to the null hypothesis $H_{02} : \text{AOR}_{l,m} = 1$, and
2. The directions of violations of H_{01} and H_{02} are the same.

These two properties also imply that

3. The ordering of the odds is the same as the ordering of the $\beta's$.

That these two properties are satisfied for the logistic model may be verified as follows. From (4), it follows that $\beta_l - \beta_m = 0$ if and only if $\mathrm{AOR}_{l,m} = 1$, which implies that Property 1 holds. Next, since $\mathrm{AOR}_{l,m}$ is monotonically increasing function of $\beta_l - \beta_m$, either of the two conditions (a) $\beta_l - \beta_m = 0(>, or <)$ and (b) $\mathrm{AOR}_{l,m} = 1(>, or <)$ implies the other, thereby showing that Property 2 holds. Also, since π_{ij} is a strictly increasing function of lp_{ij}, it follows that

$$\pi_{il} - \pi_{im} = 0(>, or <) \quad \text{if and only if} \quad lp_{il} - lp_{im} = 0(>, or <)\,. \tag{5}$$

Therefore, the ordering of the $\pi_i's$ is the same as the ordering of $\beta's$, which is the same as that of the odds associated with levels of X.

3.2 Standardized risks and SAR

The standardized risk given $X = j$ is the standardized probability defined by

$$\pi_{(S)j} = \sum_{i=1}^{s} \Pr\left(Y = 1 \mid X = j, \mathbf{Z} = z_i\right) \Pr\left(\mathbf{Z} = z_i \mid S\right),$$

where $\Pr(\mathbf{Z} = z_i \mid S)$ is the probability that $\mathbf{Z} = z_i$ in the standard population, which is denoted by S. Thus,

$$\pi_{(S)j} = \sum_{i=1}^{s} \pi_{ij}\omega_{(S)i}\,,$$

where $\omega_{(S)i} = \Pr(\mathbf{Z} = z_i \mid S)$ is nonnegative for $i = 1, 2, \ldots, s$, and $\sum_{i=1}^{s} \omega_{(S)i} = 1$.

The standardized attributable risk (SAR) for comparing level l and m of the factor X, then, is defined as

$$\mathrm{SAR}_{l,m} = \pi_{(S)l} - \pi_{(S)m} = \sum_{i=1}^{s} (\pi_{il} - \pi_{im})\, \omega_{(S)i}\,. \tag{6}$$

This SAR may be interpreted as the difference in the probabilities of the response of interest (i.e., $Y = 1$) that would be expected, if the populations of individuals with $X = l$ and $X = m$ had the same distribution into the confounding variable categories, $\mathbf{Z} = z_i$, as the standard population.

Here, we provide justification for the use of standardized risks and SARs in the *post hoc* description of significant results from the logistic regression model defined by (1) and (2). Such justification is provided by the fact that the following three properties of standardized risks hold:

1. The null hypothesis $H_{01} : \beta_l - \beta_m = 0$ is equivalent to the null hypothesis $H_{03} : \mathrm{SAR}_{l,m} = 0$, and
2. The directions of violations of H_{01} and H_{03} are the same.

These two properties imply that

3. The ordering of the standardized risks is the same as the ordering of the $\beta's$.

Properties 1 and 2 above can be verified as follows. Note that

$$lp_{il} - lp_{im} = \beta_l - \beta_m \;\forall\; i$$

from (1). It follows that $lp_{il} - lp_{im} = 0(>, or <)$ for some i implies that $lp_{il} - lp_{im} = 0(>, or <) \;\forall\; i$. Thus, from (5), we have that $\pi_{il} - \pi_{im} = 0(>, or <)$ for some i implies that $\pi_{il} - \pi_{im} = 0(>, or <) \;\forall\; i$. Therefore,

$$\beta_l - \beta_m = 0(>, or <) \quad \text{if and only if} \quad \pi_{il} - \pi_{im} = 0(>, or <) \;\forall\; i. \tag{7}$$

Properties 1 and 2 follow from the definition of $\mathrm{SAR}_{l,m}$ in (6), because the right hand side of Eq. (6) is a weighted sum of the differences $\pi_{il} - \pi_{im}$ with nonnegative weights, $\omega_{(S)i}$, where at least one $\omega_{(S)i} > 0$.

3.3 Estimated comparisons

It follows from (6) and (7) that the $\pi_{(S)l} - \pi_{(S)m}$ reflect deviations from the null hypotheses of interest, i.e., $H_0 : \beta_l - \beta_m = 0$. Thus, estimates of $\pi_{(S)l} - \pi_{(S)m}$ legitimately describe deviations from these null hypotheses and can be used for *post hoc* pairwise comparisons of the levels of X.

For a selected standard population (i.e., selected $\omega_{(S)i}$), $\mathrm{SAR}_{l,m}$ can be estimated by replacing the $\pi_{il} - \pi_{im}$ in (6) by their estimates calculated from data. One possibility is to follow the general practice of using the difference between strata specific sample proportions (i.e., $p_{il} - p_{im}$) as the estimator for $\pi_{il} - \pi_{im}$. Unfortunately, the resulting estimates are not guaranteed to be zero, positive, or negative together and the resulting estimate of SAR may not have the same sign as $\widehat{\beta_l - \beta_m}$. If, however, we use $(\hat{\pi}_{il} - \hat{\pi}_{lm})$, where the $\hat{\pi}_{ij}$ is model based estimator of π_{ij}, in the calculation of the estimated standardized risks, then the estimated comparisons have the same properties as previously shown for corresponding population parameters. That is, $\hat{\pi}_{(S)l} - \hat{\pi}_{(S)m} = 0(>, or <)$ if and only if $\widehat{\beta_l - \beta_m} = 0(>, or <)$. Thus, it is recommended that pairwise comparisons of the levels of X be calculated as the difference of standardized estimates obtained by standardizing the model-based estimates of π_{ij}.

3.4 Standard errors and confidence intervals

It is useful to provide standard errors of estimates and confidence intervals for standardized risks and SARs, even though they are used only as descriptors and are not being recommended for inference. In this section, we discuss three alternative methods for obtaining standard errors and confidence intervals: (i) Delta method; (ii) Bootstrap method; and (iii) Bayesian Markov Chain Monte Carlo (MCMC) method of posterior infernece.

3.4.1 Delta method

Let $\theta^T = (\alpha, \beta_1, \ldots, \beta_k, \delta_1, \ldots, \delta_c)$ denote the row vector of $q = k + c + 1$ parameters in the full rank representation of the logistic regression model in (1) and (2). The estimates of θ obtained from standard programs, such as PROC GENMOD with logit link in SAS (see, e.g., Stokes et al., 1995) and the LOGIT estimation command in STATA (see, e.g., Gould et al., 2003) are maximum likelihood estimates (MLEs). It is well known that the maximum likelihood estimator, $\hat{\theta}$, is asymptotically normal (see, e.g., Bender, 1999; Bura and Gastwirth, 2001) with mean θ and covariance matrix $\boldsymbol{I}^{-1}(\hat{\theta})$, where $\boldsymbol{I}^{-1}(\hat{\theta})$ is the inverse of the observed information matrix evaluated at $\theta = \hat{\theta}$. It follows that $\hat{lp}_{ij}$, calculated from $\hat{\theta}$, is the MLE of lp_{ij} for each combination of $i = 1, \ldots, s$ and $j = 1, \ldots, k$. Furthermore, since π_{ij} is a one-to-one function of lp_{ij}, the MLE of π_{ij} is

$$\hat{\pi}_{ij} = \frac{\exp\{\hat{lp}_{ij}\}}{1 + \exp\{\hat{lp}_{ij}\}}\,,$$

and the $\hat{\pi}'_{ij}s$ are jointly asymptotic normal. Therefore, the linear combinations

$$\hat{\pi}_{(S)j} = \sum_{i=1}^{s} \hat{\pi}_{ij}\omega_{(S)i} \tag{8}$$

and

$$\widehat{\mathrm{SAR}}_{l,m} = \sum_{i=1}^{s} (\hat{\pi}_{il} - \hat{\pi}_{im})\,\omega_{(S)i} \tag{9}$$

are both asymptotically normal.

The delta method can be used to approximate the standard errors of $\hat{\pi}_{(S)j}$ and $\widehat{\mathrm{SAR}}_{l,m}$ by approximating $\hat{\pi}_{(S)j}$ by a first order Taylor series about θ:

$$\hat{\pi}_{(S)j} = \pi_{(S)j} + \left[\sum_{i=1}^{s} \left(\frac{\partial \hat{\pi}_{ij}}{\partial \hat{\theta}^T}\bigg|_{\hat{\theta}=\theta}\right) \omega_{(S)i}\right] (\hat{\theta} - \theta) + O_p(1/n).$$

Therefore, an approximation to the variance of $\hat{\pi}_{(S)j}$ is

$$V_j = \left[\sum_{i=1}^{s} \left(\frac{\partial \pi_{ij}}{\partial \theta^T}\right) \omega_{(S)i}\right] [\boldsymbol{I}^{-1}(\hat{\theta})] \left[\sum_{i=1}^{s} \left(\frac{\partial \pi_{ij}}{\partial \theta^T}\right) \omega_{(S)i}\right]^T.$$

Note that $\boldsymbol{I}^{-1}(\hat{\theta})$ is the estimated large sample covariance matrix of $\hat{\theta}$ that is produced by standard software packages such as SAS and STATA. Also note that by replacing θ with its MLE, $\hat{\theta}$, V_j is estimated consistently by

$$\hat{V}_j = \left[\sum_{i=1}^{s} \left(\frac{\partial \hat{\pi}_{ij}}{\partial \hat{\theta}^T}\right) \omega_{(S)i}\right] [\boldsymbol{I}^{-1}(\hat{\theta})] \left[\sum_{i=1}^{s} \left(\frac{\partial \hat{\pi}_{ij}}{\partial \hat{\theta}^T}\right) \omega_{(S)i}\right]^T. \tag{10}$$

Thus, for the purpose of inferences, it can be assumed that

$$\hat{\pi}_{(S)j} \mathrel{\dot{\sim}} N(\pi_{(S)j}, \hat{V}_j) \tag{11}$$

when the sample size, n, is large.

Similarly, it can be shown that

$$\widehat{\mathrm{SAR}}_{l,m} = \hat{\pi}_{(S)l} - \hat{\pi}_{(S)m} \mathrel{\dot{\sim}} N(\pi_{(S)l} - \pi_{(S)m}, \hat{V}_{l,m}) \tag{12}$$

when n is large, where

$$\hat{V}_{l,m} = \hat{V}_l + \hat{V}_m - 2\widehat{\mathrm{Cov}}\,(\hat{\pi}_{(S)l}, \hat{\pi}_{(S)m}) \tag{13}$$

with $\hat{V}_l$ and $\hat{V}_m$ defined as in (10) and

$$\widehat{\mathrm{Cov}}\,(\hat{\pi}_{(S)l}, \hat{\pi}_{(S)m}) = \left[\sum_{i=1}^{s} \left(\frac{\partial \hat{\pi}_{il}}{\partial \hat{\theta}^T}\right) \omega_{(S)i}\right] [\boldsymbol{I}^{-1}(\hat{\theta})] \left[\sum_{i=1}^{s} \left(\frac{\partial \hat{\pi}_{im}}{\partial \hat{\theta}^T}\right) \omega_{(S)i}\right]^T. \tag{14}$$

Standard errors and confidence intervals for $\hat{\pi}_{(S)j}$ and $\widehat{\mathrm{SAR}}_{l,m}$ are easily derived from (10)–(14).

3.4.2 Bootstrap method

Bootstrap methods (see Efron and Tibshirani, 1993; Mooney and Duval, 1993) provide an easy way to obtain standard errors and confidence intervals for complex functions of parameters, such as $\pi_{(S)j}$ and $\mathrm{SAR}_{l,m}$. Let $\boldsymbol{a}_{ijp}^T = (Y_{ijp}, \boldsymbol{T}_{ijp}^T)$ denote the $1 \times (q+1)$ row vector of observations on the dichotomous response variable and q explanatory variables in the full-rank version of the over-parameterized model defined in (1) and (2) for the p-th subject in the j-th group and the i-th stratum. Now, let $\boldsymbol{A}$ denote the $n \times (q+1)$ matrix formed by concatenating the $\boldsymbol{a}_{ijp}^T$ row vectors. The bootstrap method of obtaining standard errors of $\hat{\pi}_{(S)j}$, say, and a $(1-\alpha) \times 100\%$ confidence interval for $\pi_{(S)j}$ is implemented as follows:

1. Sample n rows, with replacement, from the matrix $\boldsymbol{A}$ to get R resampled samples of size n.
2. Calculate $\hat{\pi}_{(S)j}$ from each sample obtained and denote the resulting estimates by $\hat{\pi}_{(S)r}$, $r = 1, 2, \ldots, R$.
3. Calculate the bootstrap standard error of $\hat{\pi}_{(S)j}$ as

$$\hat{\sigma}_j = \left[\left(\frac{1}{R-1}\right) \sum_{r=1}^{R} (\hat{\pi}_{(S)jr} - \bar{\hat{\pi}}_{(S)j\cdot})^2\right]^{1/2},$$

where

$$\bar{\hat{\pi}}_{(S)j\cdot} = \left(\frac{1}{R}\right) \sum_{r=1}^{R} \hat{\pi}_{(S)jr}.$$

4. Calculate the end-points, L and U say, of a $(1-\alpha) \times 100\%$ confidence interval for $\pi_{(S)j}$ as L = sample estimate of the $(\alpha/2)$-th percentile and U=sample estimate of the $(1-\alpha/2)$-th percentile of the population $(R = \infty)$ of resampled estimates, $\hat{\pi}_{(S)jr}$.

3.4.3 MCMC method of posterior inference

The development of software, such as the WinBugs program (Spiegelhalter et al., 2003), to generate samples of parameter vectors directly from a posterior distribution has greatly facilitated applications of Bayesian methods. Given the prior distribution and the density function of the outcome variable, WinBugs employs the Adoptive Rejection sampling algorithm for Markov Chain Monte Carlo (MCMC) sampling of parameter vectors. Arbitrarily complex functions of the parameters then can be calculated for each sample parameter vector to obtain a sample of the function's values from its posterior distribution. This sample then is used for posterior inference. Readers are referred to the article by Zuur, Garthwaite and Fryer (2002) for an in-depth discussion of practical issues in MCMC applications in Bayesian analyses.

In the current context, $\pi_{(S)j} = \sum_{i=1}^{s} \pi_{ij}\omega_{(S)i}$ and $\text{SAR}_{l,m}$ are functions of the elements of θ, the parameter vector in the logistic model. Bayesian analysis for $\pi_{(S)j}$, for example, based on MCMC samples of θ from the posterior distribution would proceed as follows:

1. Use an MCMC algorithm to generate a sample of R vectors from the posterior distribution of θ.
2. Calculate $\pi_{(S)j}$ for each sampled vector and denote the calculated values by $\pi_{(S)jr}$, $r = 1, 2, ..., R$.
3. Calculate the Bayes estimator as

$$\bar{\pi}_{(S)j} = \frac{1}{R} \sum_{r=1}^{R} \pi_{(S)jr}$$

and its standard error as

$$\hat{\sigma}_j = \left[\frac{1}{R-1} \sum_{r=1}^{R} (\pi_{(S)jr} - \bar{\pi}_{(S)j})^2\right]^{1/2}.$$

4. Calculate the end-points, L and U say, of a $(1-\alpha) \times 100\%$ credible interval for $\pi_{(S)j}$ as L = sample estimate of the $(\alpha/2)$-th percentile and U = sample estimate of the $(1-\alpha/2)$-th percentile of the posterior distribution of $\pi_{(S)j}$.

4 An Illustrative Example

The benefits of using estimated standardized risks and SARs for *post hoc* descriptions are best illustrated in an example application. In this section, we follow the analysis of a previous study (Carter et al., 2002) to demonstrate the advantages of using SARs, in addition to AORs, in making pairwise comparisons of the levels of X (i.e., lead exposure), following the observation of an overall significant effect of lead exposure. The study used a retrospective cohort design to investigate the effect and impact of high blood lead levels on educational outcomes among populations of Medicaid and non-Medicaid children in Florida.

4.1 Sample and methods

The sample analyzed consisted of 121,047 children who were born alive in Florida between September 2, 1992 and September 1, 1993 to mothers of Florida residence, entered kindergarten in a Florida Public School at 5 years of age in 1998, and had non-missing values on all variables in the analysis.

High blood lead level (BLL) was defined by BLL ≥ 10 micrograms per deciliter (μg/dl) (CDC guidelines define elevated blood lead as a single blood lead test result greater than or equal to 10 μg/dl). For illustrative purposes we focus on the academic problems outcome (i.e., retention in grade or placement in a special education class for the speech and language impaired in kindergarten) and define the response variable, Y, to be 1 if the child had academic problems and 0 otherwise. The categorical explanatory variable X has $k = 6$ levels resulting from combining the three lead exposure levels (Lead: High, if tested and BLL ≥ 10 μg/dl; Low if tested and $0 <$ BLL < 10 μg/dl; and NT if BLL $= 0$ or the child was not tested) with two levels of Medicaid participation (Medicaid: Yes, No). Because the testing program targeted children at highest risk for lead exposure, the NT group is expected to have a lower mean BLL than the Low group.

Adjustments were made for the potentially confounding effects of socio-demographic variables: mother's race (Race: Black, Other, White); mother's education (Edu: <High School, High School, >High School); mother's marital status (Mari: Not married, Married); mother's age (Age: <20, >35, 20–35); child's gender (Gend: Male, Female). Thus, $\boldsymbol{Z}$ is the vector of $c = 8$ indicator variables needed to uniquely identify the 108 possible combinations of these 5 confounding variables.

A Logistic model was fitted to estimate the effect of Lead and Lead by Medicaid interactions on Academic Problem, controlling for the other 5 categorical confounders. A significant Lead by Medicaid interaction was observed. Thus, AORs were calculated to measure the effect of lead exposure on the response within each population (Medicaid and non-Medicaid) separately. The Intersection-Union test (see Casella and Berger, 2002) was applied to test for a dose-response relationship by testing the null hypothesis

$$H_0 : \mathrm{AOR}_{\mathrm{High,\,Low}} \leq 1 \quad \text{or} \quad \mathrm{AOR}_{\mathrm{Low,\,NT}} \leq 1$$

$$(\text{or equivalently } H_0 : \pi_{(S)\mathrm{High}} - \pi_{(S)\mathrm{Low}} \leq 0 \text{ or } \pi_{(S)\mathrm{Low}} - \pi_{(S)\mathrm{NT}} \leq 0)$$

against the alternative

$$H_1 : \mathrm{AOR}_{\mathrm{High,\,Low}} > 1 \quad \text{or} \quad \mathrm{AOR}_{\mathrm{Low,\,NT}} > 1$$

$$(\text{or equivalently } H_1 : \pi_{(S)\mathrm{High}} - \pi_{(S)\mathrm{Low}} > 0 \text{ or } \pi_{(S)\mathrm{Low}} - \pi_{(S)\mathrm{NT}} > 0)$$

within each population. We also calculated the estimated standardized risks of Academic Problem ($Y = 1$) for each level of lead exposure (High, Low, NT) for the Medicaid and non-Medicaid populations using model based estimators of the probabilities of Academic Problem ($Y = 1$) for the i-th stratum, $i = 1, 2, \ldots, 108$, within the j-th level of a variable with 6 categories defined by the combinations of Lead (High, Low, NT) and Medicaid (Medicaid, non-Medicaid). The total sample was used as the standard population. Pairwise comparisons of the lead exposure groups were obtained from both the estimated AORs and SARs within, Medicaid and non-Medicaid populations separately.

Delta and Bootstrap methods were used to calculate 95% confidence intervals and 95% credible intervals were calculated by the MCMC method, using the Adaptive Rejection sampling algorithm (Gilks and Wild, 1992). A SAS, macro (see SAS Macro Language: Reference, 2000), program was written to resample the data $R = 1000$ times to obtain bootstrap estimates and confidence intervals. The WinBugs program (see WinBugs version 1.4 User manual, 2003) was used with the Adaptive Rejection sampling algorithm to generate 1000 ($= R$) samples from the posterior distribution of parameters in a full-rank logistic regression model with the prior distribution of parameters defined as independent $N(0,\ 0.001)$. After two MCMC chains of 30,000 iterations each were run the last 500 from each chain were taken producing a sample of 1000 parameter vectors from the posterior distribution. For each sampled parameter vector, standardized risks, SAR's and AOR's were calculated, and 95% credible intervals were obtained.

4.2 Results

Significant effects of Lead on Academic Problem were found in the form of a Lead by Medicaid interaction (p-value <0.0001). We summarize the findings in the Medicaid and non-Medicaid populations, separately, first by presenting AOR-based results of pairwise comparisons in Table 1 and 2.

Table 1 Estimated adjusted Odds Ratios (AORs) and 95% Confidence/Credible Intervals (CIs) for the effects of Lead exposure on Academic problem response within the Medicaid population.

Comparison	$\widehat{\text{AOR}}$	CI		
		MLE-based	Bootstrap	MCMC
NT (Reference)	1.00	–	–	–
High vs. NT	1.37	(1.20, 1.57)	(1.19, 1.57)	(1.19, 1.55)
Low vs. NT	1.19	(1.13, 1.25)	(1.13, 1.25)	(1.13, 1.25)
Low (Reference)	1.00	–	–	–
High vs. Low	1.15	(1.00, 1.32)	(0.99, 1.33)	(1.00, 1.31)

Note: Result of the Intersection-Union test for H_0: $\text{AOR}_{\text{Hish, Low}} \leq 1$ or $\text{AOR}_{\text{Low, NT}} \leq 1$ (or equivalently H_0: $\pi_{(S)\,\text{High}} - \pi_{(S)\,\text{Low}} \leq 0$ or $\pi_{(S)\,\text{Low}} - \pi_{(S)\,\text{NT}} \leq 0$) vs. H_1: $\text{AOR}_{\text{High, Low}} > 1$ and $\text{AOR}_{\text{Low, NT}} > 1$ (or equivalently H_1: $\pi_{(S)\,\text{High}} - \pi_{(S)\,\text{Low}} > 0$ and $\pi_{(S)\,\text{Low}} - \pi_{(S)\,\text{NT}} > 0$) given Medicaid population shows that there is a significant dose response effect (p-value <0.025).

Table 2 Estimated adjusted Odds Ratios (AORs) and 95% Confidence/Credible Intervals (CIs) for the effects of Lead exposure on Academic problem response within the non-Medicaid population.

Comparison	$\widehat{\text{AOR}}$	CI		
		MLE-based	Bootstrap	MCMC
NT (Reference)	1.00	–	–	–
High vs. NT	2.01	(1.44, 2.80)	(1.42, 2.74)	(1.41, 2.71)
Low vs. NT	1.42	(1.29, 1.56)	(1.28, 1.56)	(1.27, 1.56)
Low (Reference)	1.00	–	–	–
High vs. Low	1.42	(1.01, 2.00)	(0.99, 1.92)	(0.99, 1.93)

Note: Result of the Intersection-Union test for H_0: $\text{AOR}_{\text{Hish, Low}} \leq 1$ or $\text{AOR}_{\text{Low, NT}} \leq 1$ (or equivalently H_0: $\pi_{(S)\,\text{High}} - \pi_{(S)\,\text{Low}} \leq 0$ or $\pi_{(S)\,\text{Low}} - \pi_{(S)\,\text{NT}} \leq 0$) vs. H_1: $\text{AOR}_{\text{High, Low}} > 1$ and $\text{AOR}_{\text{Low, NT}} > 1$ (or equivalently H_1: $\pi_{(S)\,\text{High}} - \pi_{(S)\,\text{Low}} > 0$ and $\pi_{(S)\,\text{Low}} - \pi_{(S)\,\text{NT}} > 0$) given Medicaid population shows that there is a significant dose response effect (p-value <0.025).

There were significant dose response effects of Lead on the Academic Problem response (p-value <0.025), by the Intersection-Union test, in both the Medicaid and non-Medicaid populations.

Supplemental descriptions of results were obtained by calculating the estimated standardized risks of Academic Problem ($Y = 1$) for each level of lead exposure (High, Low, NT), by Medicaid and non-Medicaid population. The standard population used in these calculations was the entire sample. Results are presented in Figure 1 and Table 3 to illustrate the dose response relationship between lead levels and academic problems in kindergarten and differences between the Medicaid status populations. Estimates of SARs were calculated by taking differences of the standardized risk estimates. The results are given in Table 4 with standard errors and 95% confidence/credible intervals. Tests of significance of SARs are obtained from the tests of corresponding AORs in Table 1 and 2.

Table 3 Estimated Standardized Risks (SRs), Standard Errors (SEs) and 95% Confidence/Credible Intervals (CIs) for the effects of Lead exposure on Academic problem response within the Medicaid and non-Medicaid populations.

			NT	Low	High
Medicaid	Classical/Delta	$\widehat{SR}$	0.1879	0.2150	0.2389
		SE	0.0019	0.0039	0.0120
		CI	(0.1841, 0.1917)	(0.2074, 0.2226)	(0.2155, 0.2623)
	Bootstrap	$\widehat{SR}$	0.1878	0.2152	0.2388
		SE	0.0019	0.0037	0.0124
		CI	(0.1842, 0.1915)	(0.2080, 0.2222)	(0.2143, 0.2635)
	MCMC	$\widehat{SR}$	0.1878	0.2150	0.2389
		SE	0.0019	0.0040	0.0117
		CI	(0.1839,0.1913)	(0.2071, 0.2228)	(0.2156, 0.2600)
non-Medicaid	Classical/Delta	$\widehat{SR}$	0.1349	0.1797	0.2354
		SE	0.0016	0.0069	0.0293
		CI	(0.1318, 0.1380)	(0.1663, 0.1931)	(0.1779, 0.2929)
	Bootstrap	$\widehat{SR}$	0.1350	0.1800	0.2353
		SE	0.0016	0.0071	0.0289
		CI	(0.1320, 0.1384)	(0.1658, 0.1940)	(0.1819, 0.2921)
	MCMC	$\widehat{SR}$	0.1351	0.1795	0.2323
		SE	0.0016	0.0070	0.0272
		CI	(0.1320, 0.1384)	(0.1660, 0.1926)	(0.1793, 0.2911)

Figure 1 illustrates a dose-response relationship, in both Medicaid and non-Medicaid populations, in terms of the expected prevalence of academic problems given the socio-demographic mix in the standard population. For example, the expected proportion of Medicaid children with High BLL who have academic problems is about 0.24 when using the total sample as the standard population. Furthermore,

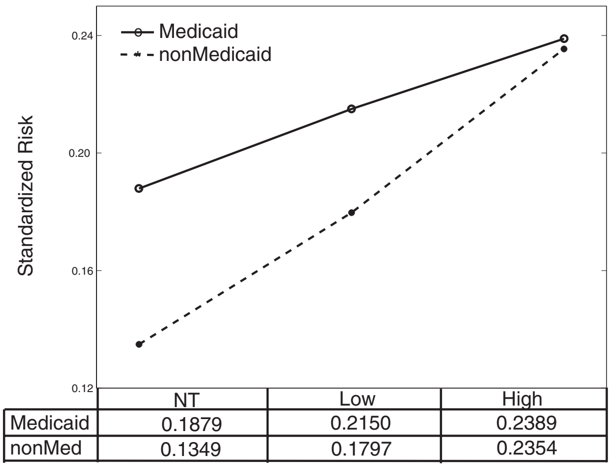

Figure 1 Estimated standardized risk of Academic Problem ($Y = 1$) for each level of Lead exposure by Medicaid and non-Medicaid populations.

Table 4 Estimated Standardized Attributable Risks (SARs), Standard Errors (SEs) and 95% Confidence/Credible Intervals (CIs) for the effects of Lead exposure on Academic problem response within the Medicaid and non-Medicaid populations.

			High vs. NT	High vs. Low	Low vs. NT
Medicaid	Classical/Delta	$\widehat{\text{SAR}}$	0.0510	0.0239	0.0271
		SE	0.0121	0.0125	0.0043
		CI	(0.0274, 0.0746)	(−0.0006, 0.0484)	(0.0188, 0.0354)
	Bootstrap	$\widehat{\text{SAR}}$	0.0510	0.0236	0.0274
		SE	0.0125	0.0129	0.0040
		CI	(0.0265, 0.0754)	(−0.0020, 0.0504)	(0.0196, 0.0351)
	MCMC	$\widehat{\text{SAR}}$	0.0511	0.0239	0.0272
		SE	0.0117	0.0122	0.0043
		CI	(0.0277, 0.0725)	(−0.0000, 0.0468)	(0.0190, 0.0353)
non-Medicaid	Classical/Delta	$\widehat{\text{SAR}}$	0.1005	0.0557	0.0448
		SE	0.0294	0.0301	0.0070
		CI	(0.0429, 0.1581)	(−0.0033, 0.1147)	(0.0310, 0.0586)
	Bootstrap	$\widehat{\text{SAR}}$	0.1003	0.0554	0.0449
		SE	0.0291	0.0300	0.0074
		CI	(0.0464, 0.1578)	(−0.0009, 0.1113)	(0.0307, 0.0590)
	MCMC	$\widehat{\text{SAR}}$	0.0972	0.0528	0.0444
		SE	0.0274	0.0283	0.0072
		CI	(0.0444, 0.1556)	(−0.0019, 0.1130)	(0.0301, 0.0585)

Medicaid children in the High Lead category have about 5.1 percentage points higher risk of academic problems than those in the NT category. This difference is statistically significant (p-value <0.05) at the 0.05 level based on the 95% confidence interval for the corresponding AOR in Tables 1 and 2. Thus, within each socio-demographic subgroup, we would expect 5 more cases of academic problems for every hundred children with high blood levels than for every hundred not tested in the Medicaid population. In the non-Medicaid population, this difference is 10 cases per hundred children. Such estimates are not obtainable from the usual *post hoc* analysis, based on AORs alone. Clearly, interpretation and communication of statistical results have been enhanced by the use of standardized risk estimates.

5 Justification for AOR and SAR in Generalized Linear Models

By using a transformation $g(\pi_{ij})$ that maps the unit interval into the real line, a more general class of binary response models can be defined by

$$g(\pi_{ij}) = lp_{ij} = \alpha + \beta_j + \sum_{g=1}^{c} \delta_g z_{ig} \,,$$

where $g(\pi_{ij})$ is invertible and is referred to as the link function. The class of models defined is the so-called generalized linear model for binary response variables (McCullagh and Nelder, 1989). There are many choices of link functions. The following four are commonly used:

1. logit: $g(\pi_{ij}) = \log\{\pi_{ij}/(1-\pi_{ij})\}$;
2. probit (or inverse Normal function): $g(\pi_{ij}) = \Phi^{-1}(\pi_{ij})$;

3. complementary log-log function: $g(\pi_{ij}) = \log\{-\log(1-\pi_{ij})\}$; and
4. log function: $g(\pi_{ij}) = \log(\pi_{ij})$.

Alternatively, the generalized linear model can be defined by

$$\pi_{ij} = F(lp_{ij}), \tag{15}$$

where F is the inverse link function and is a continuous and differentiable monotone positive valued function of lp_{ij}. That the key property

$$\beta_l - \beta_m = 0\ (>, or <)\ \forall\ i \quad \text{if and only if} \quad \mathrm{SAR}_{l,m} = 0\ (>, or <) \tag{16}$$

is satisfied under the model specified by (15) can be seen from the fact that, in our discussion of the logistic model, this result followed from the monotonicity and positivity of the logistic function of lp_{ij} in (2), independent of the form of that function.

It follows from (16) that standardized risk estimates and standardized attributable risks can be used for *post hoc* analysis and descriptors in the context of the generalized linear model for a binary response (15), as in the logistic regression model.

6 Conclusions

We have justified the use of standardized attributable risk (SAR) estimates for pairwise comparisons of the levels of a factor found to be significant in generalized linear models for binary response variables. SARs and associated standardized risk estimates provide descriptions of results that are meaningfully interpreted, while reflecting the effects measured by AORs. Furthermore, tests of significance of SARs are accomplished by testing the associated AORs. No such meaningful interpretations can be derived from pairwise analyses based either on AORs or on other results from the fitted model alone. Thus, SARs and associated standardized risk estimates are recommended as supplements to AORs in *post hoc* analysis of the results from logistic regression or other generalized linear model analysis of binary response variables.

Acknowledgement This research was funded, in part, by contract number (49102961532-12) from the Florida Agency for Health Care Administration.

References

Bender, R. (1999). Quantitative risk assessment in epidemiological studies investigating threshold effects. *Biometrical Journal* **41**, 305–319.

Bhopal, R. (2002). *Concepts of Epidemiology.* Oxford, New York.

Bura, E. and Gastwirth, J. L. (2001). The binary regression quantile plot: assessing the importance of predictors in binary regression visually. *Biometrical Journal* **43**, 5–21.

Carter, R. L., Mardon, R., Resnick, M. B., Ariet, M., Roth, J. and Feaver, E. A. (2002). The effect of lead poisoning on preschool mental development, readiness to start school, and educational disabilities. *MCHERDC Report*. University of Florida.

Casella, G. and Berger, R. L. (2002). *Statistical Inference.* Duxbury Press, California.

Cox, D. R. and Snell, E. J. (1989). *Analysis of Binary Data.* Chapman & Hall, New York.

Efron, B. and Tibshirani, R. J. (1993). *An Introduction to the Bootstrap.* Chapman & Hall, New York.

Hosmer, D. W. and Lemeshow, S. (1989). *Applied Logistic Regression.* John Wiley & Sons, New York. Flanders, W. D. and Rhodes, P. H. (1987). Large sample confidence intervals for regression standardized risk, risk ratios, and risk differences. *Journal of Chronic Diseases* **40**, 697–704.

Fletcher, R. H., Fletcher, S. W. and Wagner, E. H. (1982). *Clinical Epidemiology-the Essentials.* Williams & Wilkins, Baltimore.

Freeman, D. H. Jr. (1987). *Applied Categorical Data Analysis.* Marcel Dekker, Inc., New York.

Gilks, W. R. and Wild, P. (1992). Adaptive rejection sampling for Gibbs sampling. *Applied Statistics* **41**, 337–348.

Gould, W., Pitblado, J. and Sribney, W. (2003). *Maximum Likelihood Estimation with STATA.* STATA Press.

Greenland, S. (1991). Estimating standardized parameters from generalized linear models. *Statistics in Medicine* **10**, 1069–1074.

Greenland, S. and Holland, P. (1991). Estimating standardized risk differences from odds ratios. *Biometrics* **47**, 319–322.

Lilienfeld, D. E. (1979). The greening of epidemiology: sanitary physicians and the London epidemiological society. *Bulletin of the History of Medicine* **106**, 445–459.

McCullagh, P. and Nelder, J. A. (1989). *Generalized Linear Models.* Chapman & Hall/CRC, New York.

Mooney, C. Z. and Duval, R. D. (1993). *Bootstrapping: A Nonparametric Approach to Statistical Inference.* SAGE Publications, Inc., Newbury Park.

Neison, F. G. P. (1844). On a method recently proposed for conducting inquiries into the comparative sanitary conditions of various districts. *London Journal of the Royal Statistical Society* **7**, 40–68.

Rothman, K. J. (1998). *Modern Epidemiology.* Lippincott-Raven, Philadelphia.

SAS Institute (2000). *SAS Macro Language: Reference Version 8. 0.* SAS Institute Inc., Cary NC. Spiegelhalter, D., Thomas, A., Best, N. and Lunn, D. (2003). *WinBugs Version 1.4 User Manual.* Cambridge, UK.

Stokes, M. E., Davis, C. S. and Koch, G. G. (1995). *Categorical Data Analysis Using the SAS System.* SAS Institute Inc., Cary NC.

Uter, W. and Pfahlberg, A. (1999). The concept of attributable risk in epidemiological practice. *Biometrical Journal* **41**, 985–993.

Zuur, G., Garthwaite, P. H. and Fryer, R. J. (2002). Practical use of MCMC method: lessons from a case study. *Biometrical Journal* **44**, 433–45 5.

